

Text Mining dalam Membandingkan Metode Naïve Bayes dengan C.45 dalam Mengidentifikasi Berita Hoax pada Media SosialYola Tri Handika¹, Sarjon Defit², Gunadi Widi Nurcahyo³Universitas Putra Indonesia 'YPTK-UPI' Padang^{1,2,3}Email: hoshiyola@gmail.com¹DOI: <http://dx.doi.org/10.31869/rtj.v5i1.2855>

Abstract : Hoax news (hocus to trick) has a very big influence in disseminating information, especially in the world of social media. News has an important impact on social and political conditions, and news can move the economy of a country. For this reason, it is necessary to have an analysis to classify hoax news and not hoaxes, and have high accuracy in classifying the news. In this study, two methods were used as a comparison in achieving high accuracy, namely the Naïve Bayes method which is famous for having high accuracy in classification with little data, and the C.45 method which can minimize noise in the data. The data used are 300 articles with 10 topics which contain hoax and non-hoax news. The data is obtained from the internet through social media, such as Twitter, Instagram and Facebook. Testing using the Naïve Bayes method has a higher accuracy than the C.45 method. The amount of data used has a major influence on the test results, if more data enters the training stage, then this study will have higher accuracy. However, the results of this test can be recommended to increase accuracy in the construction of a hoax news detection system.

Keywords: text mining, naïve bayes, C.45, hoax, social media

Abstrak: Berita Hoaks (*hocus to trick*) memiliki pengaruh yang sangat besar dalam penyebaran informasi, khususnya dalam dunia media soSial. Berita memiliki dampak yang penting dalam kondisi sosial maupun politik, dan berita dapat menggerakkan perekonomian pada suatu negara. Untuk itu perlu adanya analisa untuk mengklasifikasikan berita *hoax* dan bukan *hoax*, serta memiliki keakuratan yang tinggi dalam pengklasifikasian berita tersebut. Pada penelitian ini digunakan dua metode sebagai perbandingan dalam pencapaian akurasi yang tinggi, yaitu metode *Naïve Bayes* yang terkenal memiliki akurasi yang tinggi dalam klasifikasi dengan data yang sedikit, serta metode C.45 yang dapat meminimalkan *noise* pada data. Data yang digunakan berjumlah 300 artikel dengan 10 topik yang didalamnya terkandung berita hoaks dan non hoaks. Data tersebut diperoleh dari internet melalui media sosial, seperti twitter, instagram dan facebook. Pengujian dengan menggunakan metode *Naïve Bayes* memiliki akurasi yang lebih tinggi disbanding dengan metode C.45. Jumlah data yang digunakan memiliki pengaruh besar terhadap hasil pengujian, jika data yang masuk ke tahap *training* lebih banyak, maka penelitian ini akan memiliki keakurasian yang lebih tinggi. Namun, hasil pengujian ini dapat direkomendasikan untuk meningkatkan keakurasian dalam pembangunan sistem pendeteksian berita *hoax*.

Kata kunci: text mining, naïve bayes, C.45, hoax, media sosial

PENDAHULUAN

Knowledge Discovery in Database (KDD) merupakan proses mengidentifikasi pattern atau pola didalam data yang benar, unik, berguna dan dapat dimengerti. KDD memiliki beberapa step yaitu *Selection, Preprocessing, Transformation, Data Mining dan Interpretation/Evaluation*. Pada dasarnya *Text Mining* memiliki pengolahan data yang sama dengan *Data Mining*, yang membedakannya adalah jenis data yang digunakan, jika *Data Mining* mengelola data yang terstruktur, sedangkan *Text Mining* mengelola data yang tidak terstruktur. Pattern yang dihasilkan KDD dapat menghasilkan

beberapa jenis proses, di antaranya *Prediction, Classification, Clustering dan Association* [1].

Dalam penerapan *Text Mining*, pengolahan data dilakukan dalam jumlah yang banyak sebagai bahan pembelajaran dalam melakukan pelatihan untuk pengklasifikasian text dalam suatu klasifikasi tertentu seperti klasifikasi pendapat yang bersifat biasa atau sarkasme pada Media Sosial Twitter [2]. Pendeteksian berita *hoax* ini dapat diwujudkan dalam suatu *software* yang dapat dimanfaatkan untuk berbagai kepentingan yang tentunya dapat membantu dalam mengatasi misinformasi yang akan merugikan masyarakat [3]. Selain pada dunia maya,

penggunaan *Text Mining* juga dilakukan dalam berbagai bidang, seperti bidang kesehatan yang dapat memanfaatkan *Text Mining* dalam korelasi antara medis dan diagnosa pada data rekam medis pasien [4]. Metode Naïve Bayes juga digunakan dalam pengklasifikasi gambar promosi dengan non promosi dalam media sosial, sehingga gambar yang berisi konten promosi tidak terkubur oleh konten yang bersifat non promosi [5].

Penelitian terdahulu mengangkat permasalahan klasifikasi berita hoax menggunakan metode Gaussian Naïve Bayes dan setelah melalui percobaan telah didapatkan hasil bahwa klasifikasi meningkat dengan penggunaan Naïve Bayes Classifier Bernouli dibandingkan dengan Gaussian Naïve Bayes [6]. Selain itu terdapat penelitian mengenai pemilihan model dalam pendeteksian berita hoax, yaitu penggunaan BERT dan model terlatih serupa memiliki kinerja terbaik untuk berita palsu, terutama set data yang sangat kecil [7].

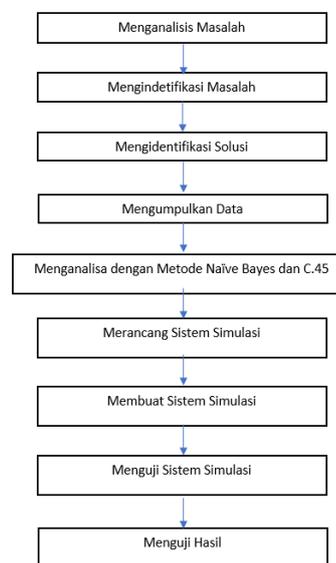
Dalam penerapan *Text Mining*, pengolahan data dilakukan dalam jumlah yang banyak sebagai bahan pembelajaran dalam melakukan pelatihan untuk pengklasifikasian text dalam suatu klasifikasi tertentu seperti klasifikasi pendapat yang bersifat biasa atau sarkasme pada Media Sosial Twitter. Pendeteksian berita *hoax* ini dapat diwujudkan dalam suatu *software* yang dapat dimanfaatkan untuk berbagai kepentingan yang tentunya dapat membantu dalam mengatasi misinformasi yang akan merugikan masyarakat [3]. Selain pada dunia maya, penggunaan *Text Mining* juga dilakukan dalam berbagai bidang, seperti bidang kesehatan yang dapat memanfaatkan *Text Mining* dalam korelasi antara medis dan diagnosa pada data rekam medis pasien [4]. Metode Naïve Bayes juga digunakan dalam pengklasifikasi gambar promosi dengan non promosi dalam media sosial, sehingga gambar yang berisi konten promosi tidak terkubur oleh konten yang bersifat non promosi [5].

Berdasarkan masalah di atas, maka pada penelitian ini akan dilakukan perbandingan penggunaan Metode Naïve Bayes dan C.45 dalam klasifikasi berita *hoax*, perbandingan tersebut akan terlihat pada tingkat akurasi dari masing-masing metode. Selain melakukan klasifikasi berita *hoax*, pada penelitian ini juga akan mencari hubungan metode seleksi fitur

terhadap akurasi klasifikasi. Sehingga akan memperlihatkan apakah penambahan metode seleksi fitur dapat meningkatkan akurasi atau tidak.

METODOLOGI PENELITIAN

Metodologi adalah sebuah tata cara yang menentukan proses penelusuran apa yang ingin digunakan. Metodologi Penelitian merupakan sekumpulan kegiatan, peraturan serta prosedur yang dipakai oleh peneliti suatu disiplin ilmu. Dalam hal ini peneliti akan melaksanakan metodologi penelitian pada bidang ilmu Teknologi dan penerapannya yaitu penggunaan metode *Naïve Bayes* dan C.45 pada klasifikasi berita *hoax*.



Gambar 1. Metodologi Penelitian

Mengidentifikasi Masalah

Tahapan ini merupakan tahapan awal saat melakukan penelitian. Tujuan awal dari suatu penelitian adalah mencari solusi dari suatu masalah. Sehingga peneliti harus mengidentifikasi masalah yang akan diselesaikan dalam penelitian ini, agar didapatkan suatu solusi yang tepat dan terukur.

Menganalisa Masalah

Dalam tahapan menganalisa masalah, peneliti harus mempelajari dan memahami masalah tersebut sehingga didapatkan ruang lingkup serta batasan masalah yang akan menjadi dasar dari penelitian ini. Pada penelitian yang akan dilakukan, peneliti akan meneliti mengenai dampak dari berita *hoax* yang tersebar di media sosial yang mengakibatkan keributan di tengah masyarakat. Selanjutnya peneliti akan memilih beberapa topic yang akan dijadikan

sampel dalam penelitian tentang berita hoax yang beredar di media sosial.

Mengidentifikasi Solusi

Untuk mendapatkan solusi dari permasalahan yang terjadi, maka akan diidentifikasi solusi-solusi yang akan dilakukan untuk mengatasi permasalahan tersebut. Selanjutnya solusi tersebut akan dipilih untuk menjadi solusi yang terbaik. Pada penelitian ini, solusi yang diambil dengan menerapkan Metode Naïve Bayes dan C.45, agar dapat membedakan antara berita hoax dan bukan hoax.

Mengumpulkan Data

Pada tahap pengumpulan data ini digunakan teknik studi pustaka dengan mengumpulkan 30 topik yang banyak beredar pada media sosial yang menimbulkan keresahan di tengah masyarakat. Media sosial yang digunakan adalah Twitter, Facebook dan Instagram. Selanjutnya, dari masing-masing topik akan dikumpulkan 10 berita yang terdiri dari berita hoax dan non hoax, sehingga data yang digunakan berjumlah 300 data. Cara pencarian berita dilakukan dengan menginputkan kata kunci pada *search engine* yang terdapat pada masing-masing media sosial. Selain itu studi pustaka juga dilakukan pada Portal Hoax yang tersedia pada Website Kementerian Komunikasi dan Informatika RI dimana pada portal tersebut terdapat berita-berita hoax yang telah terverifikasi oleh tenaga ahli. Data-data tersebut akan dijadikan sampel dalam penelitian klasifikasi berita *hoax*.

Menganalisa dengan Metode Naïve Bayes

Pada tahapan ini akan dilakukan analisa menggunakan metode Naïve Bayes dan C.45 dengan menganalisis variabel-variabel yang akan digunakan dan perhitungan untuk mengukur tingkat akurasi dari masing-masing metode. Perhitungan dilakukan dengan menggunakan *Microsoft Excel* untuk simulasi perhitungan data.

Merancang Sistem Simulasi

Pada tahapan ini akan dirancang sistem simulasi yang akan digunakan. Penelitian ini menggunakan *Edraw* untuk merancang alur sistem simulasi dan *Rapidminer* dalam membuat sistem simulasi. Selanjutnya akan dirancang modul-modul yang akan digunakan dalam penerapan metode serta tahapan-tahapan yang dilakukan sebelum menerapkan metode yang digunakan.

Membuat Sistem Simulasi

Pada tahapan ini dilakukan pembuatan sistem simulasi dengan menggunakan *Rapidminer*. Output dari sistem ini adalah klasifikasi berita hoax dan non hoax beserta dengan akurasi. Serta simulasi lain yang dilakukan dengan mengurangi beberapa proses sesuai dengan tujuan penelitian.

Menguji Sistem Simulasi

Setelah sistem simulasi dibuat, maka akan dilakukan pengujian terhadap sistem. Baik itu dari segi data, variabel, proses maupun output yang dihasilkan. Pada penelitian ini akan dilakukan pengujian terhadap nilai akurasi terhadap penggunaan Metode Naïve Bayes dan Metode C.45. Kedua nilai tersebut akan dibandingkan, sehingga akan didapatkan kesimpulan metode yang paling tepat digunakan dalam penyelesaian masalah ini.

Menguji Hasil

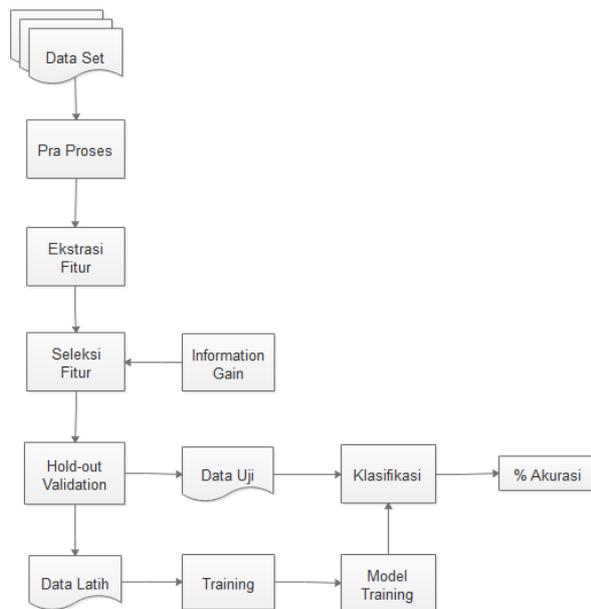
Pada tahapan ini peneliti akan melakukan pengujian terhadap sistem yang telah dirancang, dan terdapat beberapan langkah dalam pengujian hasil penelitian adalah sebagai berikut :

1. Pengumpulan data berita hoax dan non hoax dengan topic yang telah ditentukan pada media sosial
2. Melakukan analisa terhadap data yang telah dikumpulkan atau melakukan pra proses pada data
3. Melakukan simulasi terhadap data dengan menerapkan Metode Naïve Bayes dan C.45 dengan menggunakan *Rapidminer*.

HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan tahap analisis sistem yang bertujuan untuk menganalisa setiap proses yang dilakukan oleh sistem, di mana pada penelitian ini akan melakukan klasifikasi artikel *hoax* dan bukan *hoax* seperti yang telah dijelaskan pada bab sebelumnya. Selain itu juga dijelaskan perancangan sistem klasifikasi artikel sehingga akan memperoleh ciri atau fitur yang akan menjadi pembeda antara artikel *hoax* dan bukan *hoax*.

Berikut ini dijelaskan pada Gambar 4.1 mengenai rancangan tahapan klasifikasi yang menggambarkan bagaimana proses simulasi klasifikasi berita *hoax* dan bukan *hoax* akan dilakukan.



Gambar 2. Tahapan Sistem

Data

Data yang digunakan pada penelitian ini adalah kumpulan artikel atau berita yang diperoleh dari media sosial yang terdiri dari berita hoax dan bukan hoax. Kegiatan pertama yang dilakukan dengan mencari topik-topik artikel hoax yang paling populer menggunakan kueri pada search engine, menetapkan daftar topic artikel hoax yang digunakan kemudian dilanjutkan dengan mengumpulkan artikel yang terkait dengan topik tersebut.

Berdasarkan kegiatan tersebut, ditetapkan 15 topik berita hoax yang digunakan dalam penelitian ini. Data yang digunakan terdiri dari 300 berita hoax di mana berita tersebut mengandung salah satu dari topic yang telah ditentukan. Berita tersebut terdiri dari 150 berita hoax dan 150 berita bukan hoax. Pelabelan berita atau artikel dilakukan secara manual dengan merujuk pada referensi kanal berita yang terpercaya. Tabel berikut berisi daftar topic berita hoax yang tersebar dan terkenal di internet yang digunakan dalam penelitian ini.

Tabel 1. Topik Berita Hoax yang digunakan

No	Topik Berita Hoax
1	Vaksin Sinovac
2	Vaksin AstraZeneca
3	Pemberlakuan Pembatasan Kegiatan Masyarakat
4	Penerimaan CPNS Tahun 2021
5	Covid-19
6	Penerimaan Peserta Didik Baru (PPDB Online)

- 7 Vaksin Nusantara
- 8 Paket Kuota Kemendikbud
- 9 Manfaat Minum Air Putih
- 10 Obat Covid
- 11 Kartu Prakerja
- 12 PSBB
- 13 Stimulus PLN
- 14 Omnibus Law
- 15 Gejala Covid

Case Folding pada Artikel

Tahap pertama pada *text processing* adalah melakukan proses *Case Folding* yang dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar yang biasanya huruf kecil atau *lowercase*. *Case Folding* adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima oleh sistem, sedangkan komponen-komponen selain huruf dianggap sebagai delimeter dan akan dihapus dari dokumen.

Hasil akhir dari *case folding* adalah dokumen yang telah distandarkan sehingga sistem dapat menerima informasi yang jelas dari dokumen tersebut. Selanjutnya dokumen tersebut akan menjadi input untuk tahapan tokenisasi. Tabel 4.3 merupakan contoh artikel yang telah melalui proses *Case Folding*.

Tokenisasi pada Artikel

Tahapan kedua dalam *text processing* adalah proses tokenisasi. Yaitu proses pemotongan/pemisahan kalimat menjadi kata. Artikel yang terdiri dari beberapa kalimat akan dipisahkan menjadi kata-kata penyusunnya. Proses tokenisasi ini menggunakan fungsi tokenisasi dalam simulasinya. Sehingga akan didapatkan *string* yang merupakan hasil tokenisasi. Hasil pada proses ini kemudian akan disimpan untuk menjadi inputan pada proses selanjutnya. Tabel 4.4 merupakan contoh artikel sebelum melalui proses tokenisasi dan sesudah dilakukan tokenisasi.

Stopwords pada Artikel

Tahapan Stopwords merupakan tahapan yang digunakan untuk mengambil kata-kata penting dari hasil tokenisasi. Proses ini dapat menggunakan algoritma Stoplist (membuang kata kurang penting) atau Wordlist (menyimpan kata penting). Stoplist/Stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam suatu kalimat. Dapat diartikan bahwa kata pada stopwords tersebut

merupakan kata yang tidak memiliki arti jika berdiri sendiri. Contoh stopwords adalah “yang”, “dan”, “di”, “adalah”, “dari”, “ke” dan seterusnya. Data stopwords tidak memiliki standar, dapat disusun sendiri oleh peneliti atau dapat merujuk pada penelitian sebelumnya.

Stemming pada Artikel

Pada tahapan ini dokumen yang digunakan merupakan hasil dari proses stopwords yang telah dilakukans sebelumnya. Proses *Stemming* merupakan proses pemotongan imbuhan yang melekat pada kata yang terdapat pada artikel. Kata-kata pada artikel akan dikembalikan ke kata dasarnya. Proses ini bertujuan untuk mengurangi variasi kata yang sebenarnya memiliki kata dasar yang sama. Tujuan *Stemming* diperlukan untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapat imbuhan yang berbeda. Selanjutnya data tersebut akan masuk ke tahap ekstraksi fitur.

Proses pada proses stemming ini dimulai dengan melakukan pemeriksaan dan pengecekan apakah variabel penampung mengandung daftar kata dasar. Daftar kata dasar sendiri disimpan dalam sebuah file. Jika kata merupakan kata dasar, maka kata dalam artikel tidak mengalami apa-apa. Sedangkan jika bukan termasuk kedalam kata dasar, maka fungsi mengecek apakah kata mengandung imbuhan *suffiks* = {"i", "an", "ku", "mu", "ke", "se", "te", "be", "nya"} dan seterusnya}, *prefix* = {"di", "ke", "se", "te", "be", "ber", "peng", "bel"} dan seterusnya} dan *pronoun* = {"ku", "mu", "nya"}.

Ekstraksi Fitur

Setelah data melalui proses pembersihan, selanjutnya data akan melakukan ekstraksi fitur pada artikel *hoax* dan bukan *hoax*. Pada penelitian ini terdapat tiga jenis fitur yang digunakan yaitu : unigram, bigram dan trigram. Model fitur tersebut dipresentasikan dalam sebuah dataset. Tujuan penggunaan teknik N-gram ini adalah untuk memecah kata karena pada kalimat Bahasa Indonesia, suatu frasa memiliki kemungkinan terdiri dari lebih

dari satu kata. Tabel 4.6 merupakan contoh penggunaan unigram, bigram dan trigram

Proses Ekstraksi Fitur pada text mining yaitu pengambilan fitur dari suatu artikel atau dokumen yang menjadikan ciri suatu artikel tersebut *hoax* atau bukan. Berikut ini merupakan contoh pengambilan fitur dari suatu artikel. Sesuai pembahasan pada sub bab sebelumnya, data ini telah melalui proses Pra Proses terlebih dahulu.

Seleksi Fitur dengan *Information Gain*

Berikut ini merupakan contoh perhitungan *Information Gain* :

1.1 Tabel 3. Matriks Kata

		k	er	u	Ma	y	bu	b	A	su	p
K	Fit	ka	a	nu	a	n	k	u	sa		
*	ur	n	m	h	yi	di	a	k	pi		
1	D1	1	1	1	1	0	0	0	0		
0	D2	1	1	0	0	1	1	1	0		
1	D3	1	1	0	0	0	1	0	1		

*1 : Hoax, 0 : Bukan Hoax

Berdasarkan rumus dalam teknik *Information Gain* terdapat beberapa tahapan dalam perhitungan bobot *Information Gain* adalah sebagai berikut :

1. Cari nilai Entropy sebelum pemisahan sesuai dengan Rumus

$$- 'bunuh' = - 2/3 \text{Log}_2 2/3 + (-1/3 \text{Log}_2 1/3)$$

$$= 0,528320834 + 0,389975$$

$$= 0,918$$
2. Cari nilai Entropy setelah pemisahan sesuai dengan Rumus (8)

$$- 'bunuh' = 2/3 (-1/1 \text{Log}_2 1/1) + (-0/1 \text{Log}_2 0/1) + 1/3 (-1/2 \text{Log}_2 1/2)$$

$$+ (-1/2 \text{Log}_2 1/2)$$

$$= 0,667$$
3. Cari nilai *Information Gain* sesuai dengan Rumus (9)

$$\text{Gain}(S,A) = \text{Entropy}(S) - \text{Entropy}(S,A)$$

$$= 0,918 - 0,667$$

$$= 0,252$$

Klasifikasi dengan Naïve Bayes

Berdasarkan teorema Naïve Bayes yang telah dijelaskan pada bab sebelumnya, berikut ini merupakan contoh perhitungan Algoritma Naïve Bayes pada klasifikasi *Text Mining*. Data yang digunakan adalah hasil keluaran

dari seleksi fitur yang dibuatkan dalam bentuk matriks yang dapat dilihat pada Tabel 2.

melalui praproses, ekstraksi fitur dan seleksi fitur.

Tabel 2. Hasil Ekstraksi Fitur

Fitur	andi	Kerupuk	Bayi	bunuh	Suka	Sapi	makan	Bayam	Klasifikasi
D1	0	0	1	1	0	0	1	1	Hoax
D3	0	0	0	0	1	1	1	1	Hoax
D2	1	1	0	0	1	0	1	1	Bukan Hoax

Data dikelompokkan berdasarkan kategori agar memudahkan dalam perhitungan. Selanjutnya akan dihitung beberapa komponen yang akan menjadi parameter dalam menghitung probabilitas.

Tabel 4. Komponen Perhitungan Probabilitas

Kategori	Jumlah Dokumen di setiap kategori	Jumlah Dokumen Keseluruhan	P(kategori)	Jumlah kata pada setiap kategori
Hoax	2	3	2,5	16
Bukan hoax	0	3	1,5	8

Jumlah Kata Unik : 8'

Selanjutnya dilakukan penghitungan probabilitas tiap kata pada setiap artikel yang digambarkan pada Tabel berikut :

Tabel 5. Probabilitas Tiap Kata

	Hoax	Bukan Hoax
Andi	0,125	0,125
Kerupuk	0,125	-
Bayi	-	-
Bunuh	-	-
Suka	-	-
Sapi	-	0,125
Makan	-	-
Bayam	-	-

Jika telah mendapatkan hasil probabilitas tiap kata, maka untuk tahap selanjutnya yaitu menghitung probabilitas tiap kata pada artikel yang akan digunakan, baik itu data latih maupun data uji. Berikut contoh yang akan digunakan untuk menguji probabilitas dari suatu artikel yang belum ada labelnya. Contoh Artikel : Andi suka makan bayi. Maka akan ditampilkan dalam bentuk matriks yang telah

Tabel 6. Probabilitas Data Testing

	andi	Suka	Makan	bayi	Total Probabilitas	Hasil
Hoax	0	0,125	0,125	0,125	2,875	2,875
Bukan Hoax	0,13	0,125	0,125	0,125	2,00	

Sehingga dapat disimpulkan bahwa artikel yang berisikan tulisan "Andi suka makan bayi" adalah *hoax* dengan tingkat probabilitas 2,875.

Klasifikasi dengan C.45

Pada tahap ini akan dilakukan klasifikasi menggunakan Algoritma C.45, di mana algoritma ini dikenal dengan algoritma klasifikasi yang cukup efektif dengan meminimalisir *noise*. Klasifikasi ini menggunakan teknik *Hold-out Validation* dengan membagi data set awal menjadi data latih dan data uji. Sehingga sistem tidak hanya berfokus pada data yang digunakan saat melakukan simulasi.

Tabel 7. Perhitungan Nilai Gain

Kode	Jumlah Kasus (S1)	Hoax (S1)	Bukan Hoax (S2)	Entropy	Gain
1	Klasifikasi	3	2	1	0,91 8295 834
	Andi				0,91 8295 834
		1	1	0	1
		0	2	2	0
	kerupuk				0,91 8295

					834
	1	1	0	1	0
	0	2	2	0	0
Bayi					0,25 1629 167
	1	1	1	0	0
	0	2	1	1	1
bunuh					0,25 1629 167
	1	1	1	0	0
	0	2	1	1	1
suka					0,25 1629 167
	1	2	1	1	1
	0	1	1	0	0
sapi					0,25 1629 167
	1	1	1	0	0
	0	2	1	1	1
makan					0
					0,91 8295
	1			1	834
	0	0	0	0	0
bayam					0
					0,91 8295
	1			1	834
	0	0	0	0	0

Pada tabel di atas, *entropy* yang telah bernilai 0 maka proses akan berhenti dan akan dilihat nilai *gain* yang paling tinggi untuk dijadikan akar selanjutnya. Proses partisi pohon keputusan akan berhenti saat semua tupel berada pada node N mendapat kelas yang sama dan atau tidak ada atribut di dalam tupel yang dipartisi lagi dan atau tidak ada tupel pada cabang yang kosong

PENUTUP

Pada penelitian ini dapat disimpulkan bahwa penggunaan Metode Naïve Bayes dan C.45 telah terbukti baik dalam proses klasifikasi. Dalam hal ini metode Naïve Bayes memiliki tingkat akurasi yang lebih tinggi dari Metode C.45 dalam klasifikasi berita *hoax*.

Algoritma yang terbaik dalam proses klasifikasi berita *hoax* dengan menggunakan sistem operasi seleksi fitur.

DAFTAR RUJUKAN

- [1] Kim, H., Soibelman, L., & Grobler, F. (2008). Factor selection for delay analysis using Knowledge Discovery in Databases. *Automation in Construction*, 17(5), 550–560. doi:10.1016/j.autcon.2007.10.001
- [2] Khairunnisa, S., Adiwijaya, A., & Faraby, S. A. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406. doi:10.30865/mib.v5i2.2835
- [3] Englmeier, K. (2021). The Role of Text Mining in Mitigating the Threats from Fake News and Misinformation in Times of Corona. *Procedia Computer Science*, 181, 149–156. doi:10.1016/j.procs.2021.01.115
- [4] Chintalapudi, N., Battineni, G., Canio, M. D., Sagaro, G. G., & Amenta, F. (2021). Text mining with sentiment analysis on seafarers' medical documents. *International Journal of Information Management Data Insights*, 1(1), 100005. doi:10.1016/j.jjime.2020.100005
- [5] Hubert, Phoenix, P., Sudaryono, R., & Suhartono, D. (2021). Classifying Promotion Images Using Optical Character Recognition and Naïve Bayes Classifier. *Procedia Computer Science*, 179, 498–506. doi:10.1016/j.procs.2021.01.033
- [6] Singh, M., Wasim Bhatt, M., Bedi, H. S., & Mishra, U. (2020). Performance of Bernoulli's naive bayes classifier in the detection of fake news. *Materials Today: Proceedings*. doi:10.1016/j.matpr.2020.10.896
- [7] Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4, 100032. doi:10.1016/j.mlwa.2021.100032
- [8] Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Da*

- n Ilmu Komputer, 5(4), 427. doi:10.25126/jtiik.201854773
- [9] Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1). doi:10.31294/jki.v7i1.5740
- [10] Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455. doi:10.25126/jtiik.201854803
- [11] Irena, B., & Erwin Budi Setiawan. (2020). Fake News (Hoax) Identification on Social Media Twitter using Decision Tree C4.5 Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 711–716. doi:10.29207/resti.v4i4.2125
- [12] Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4, 100032. doi:10.1016/j.mlwa.2021.100032
- [13] Prasetyo, A. B., Isnanto, R. R., Eridani, D., Soetrisno, Y. A. A., Arfan, M., & Sofwan, A. (2017). Hoax detection system on Indonesian news sites based on text classification using SVM and SGD. *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. doi:10.1109/icitacee.2017.8257673
- [14] Negara, A. B. P., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(3), 599. doi:10.25126/jtiik.2020.711947
- [15] Pujiyanto, U., & Ristanti, P. Y. (2019). Perbandingan kinerja metode C4.5 dan Naive Bayes dalam klasifikasi artikel jurnal PGSD berdasarkan mata pelajaran. *TEKNO*, 29(1), 50. doi:10.17977/um034v29i1p50-67
- [16] Kim, H., Soibelman, L., & Grobler, F. (2008). Factor selection for delay analysis using Knowledge Discovery in Databases. *Automation in Construction*, 17(5), 550–560. doi:10.1016/j.autcon.2007.10.001
- [17] Rahutomo, F., Pratiwi, I. Y. R., & Ramadhani, D. M. (2019). Eksperimen Naive Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia. *JURNAL PENELITIAN KOMUNIKASI DAN OPINI PUBLIK*, 23(1). doi:10.33299/jpkop.23.1.1805
- [18] Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 427. doi:10.25126/jtiik.201854773
- [19] Dyo fatra Abdul Hadiy, Hayatin, N. H., & Aditya, C. S. K. (2020). Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia. *Jurnal Repositor*, 2(11), 1562. doi:10.22219/repositor.v2i11.933
- [20] W, B., Riski, I., Dwi, K., Nooraeni, R., Sihahan, T., & Dhea, Y. (2019). Analisis Text Mining dari Cuitan Twitter Mengenai Infrastruktur di Indonesia dengan Metode Klasifikasi Naive Bayes. *EIGEN MATHEMATICS JOURNAL*, 1(2), 92. doi:10.29303/emj.v1i2.36